PARLOMA – A Novel Human-Robot Interaction System for Deaf-blind Remote Communication

International Journal of Advanced Robotic Systems

OPEN ACCESS
ARTICLE

PARLOMA – A Novel Human-Robot Interaction System for Deaf-Blind Remote Communication

Regular Paper

Ludovico Orlando Russo[1*], Giuseppe Airò Farulla[1], Daniele Pianu[2], Alice Rita Salgarella[3], Marco Controzzi[3], Christian Cipriani[3], Calogero Maria Oddo[3], Carlo Geraci[4], Stefano Rosa[1] and Marco Indaco[1]

1 Politecnico di Torino, Department of Control and Computer Engineering, Italy
2 Institute of Electronics, Computer and Telecommunication Engineering, National Research Council, Italy
3 The BioRobotics Institute, Scuola Superiore Sant'Anna, Pisa, Italy
4 CNRS, Institute Jean-Nicod, Paris, France
* Corresponding author(s) E-mail: ludovico.russo@polito.it

Received 02 October 2014; Accepted 22 December 2014

DOI: 10.5772/60416

Abstract

Deaf-blindness forces people to live in isolation. At present, there is no existing technological solution enabling two (or many) deaf-blind people to communicate remotely among themselves in tactile Sign Language (t-SL). When resorting to t-SL, deaf-blind people can communicate only with people physically present in the same place, because they are required to reciprocally explore their hands to exchange messages. We present a preliminary version of PARLOMA, a novel system to enable remote communication between deaf-blind persons. It is composed of a low-cost depth sensor as the only input device, paired with a robotic hand as the output device. Essentially, any user can perform hand-shapes in front of the depth sensor. The system is able to recognize a set of hand-shapes that are sent over the web and reproduced by an anthropomorphic robotic hand. PARLOMA can work as a "telephone" for deaf-blind people. Hence, it will dramatically improve the quality of life of deaf-blind persons. PARLOMA has been presented and supported by the main Italian deaf-blind association, Lega del Filo d'Oro. End users are involved in the design phase.

Keywords Human-Robot Interaction, Hand Gesture Recognition, Haptic Interface, Assistive Robotics

1. Introduction

Recent technological advances in modern low-cost sensing technologies have pushed the research on Human-Robot Interaction (HRI) towards the development of natural and intuitive interaction techniques. Usually, human beings interact with machines using a mouse, a keyboard and a joystick. Unfortunately, these devices inhibit interaction in applications that require human-robot collaboration [1]. Elderly and disabled people may experience even more serious issues in using these interaction devices.

In response to this, researchers have investigated interfaces based on the natural modalities that people already use to

interact with each other. The proposed interaction interfaces are intuitive for users, and do not require learning new modalities of interaction. For instance, hand gestures are a natural part of human interaction with both machines and other humans. They are an intuitive and simple way to transmit information and commands (such as zoom in or out, drag and drop). Hand Gesture Recognition (HGR) has gained momentum within the field of HRI, becoming an important research topic. However marker-less and robust solutions for the recognition of complex gestures in real-time are still lacking [2].

The recent availability of innovative, low-cost, off-the-shelf input devices, such as structured light cameras [3], have enabled access to a finer-grained set of input data that can be exploited to enhance HGR algorithms. RGB-D cameras have fuelled the development of innovative solutions. Moreover, the impressive development of General-Purpose computing on Graphics Processing Units (GPGPUs), paired with a consolidated modern programming framework, enable higher throughput/performance on highly parallel problems, such as the image-processing tasks performed during HGR. Several approaches have been presented in the literature for HGR [4] - [8] with solutions based on different mathematical models, input cameras (e.g., RGB, depth cameras and multi-cameras systems) and algorithms. However, a quick, robust, natural and intuitive solution has yet to be found, as existing approaches very often require an intensive tuning phase, the usage of coloured or sensitized gloves, or a working framework which embeds more than one imaging sensor.

Human interaction widely uses hand gestures, and in special elements of the population gestures also serve to develop a natural language. This is the case for deaf communities who have developed Sign Languages (SLs) to serve their communication needs. SLs are independently-developed natural languages and, despite having different modalities, they exhibit more or less the same level of complexity as spoken languages [9]. Tactile Sign Language (t-SL) is the adaptation of any SL made by deaf-blind signers in order to receive linguistic messages. Deaf-blind people cannot hear or see. They cannot access information by resorting to vocal modalities or visual-based SLs. Although it cannot be technically defined a natural language, t-SL incorporates many of the features that natural languages have [10]. t-SL messages are not perceived by the visual channel, as in standard SL exchanges, but rather by tactile exploration of the hands of the person who is signing. This is required when the communication is addressed to a deaf-blind signer, who may reply in t-SL if the interlocutor is deaf-blind himself, or else in the standard SL if the interlocutor is sighted (i.e., deaf-blind signers may produce either SL or t-SL). In both SL and t-SL interactions, real-time constraints play an important role as they influence the naturalness desired for comprehensive communication. At the same time, the needs of the target population are such that HGR systems which use markers are perceived as cumbersome and inhibiting natural interaction, and thus they are not used. This population will benefit from unaided single camera-based HGR systems, where no marker and no tuning phase is required and calibration tasks are simplified. In fact, any extensive initialization task represents a barrier for users who are not at ease with technology, especially if they experience severe disabilities like deaf-blindness.

PARLOMA is developed for people who use t-SL as their main communication system - in particular, persons affected by Usher Syndrome type 1 (namely, they are deaf from birth and progressively lose their sight during adolescence), deaf individuals who had SL as their first language before becoming blind, and individuals who are born deaf-blind. Deaf-blind signers prefer to use t-SL over other spoken language-based systems, like the Braille alphabet, because t-SL makes communication more natural and effective. Despite being more efficient than other communication systems, t-SL forces pairwise communication and requires the two interlocutors to be in the same location (otherwise tactile exploration cannot happen), causing severe limitations to communication among deaf-blind people.

Remote communication systems will dramatically improve social inclusion and participation in active society for deaf-blind persons, enabling a level of access to information and to interaction with the community. Moreover, they will heavily influence the perception that deaf-blind people have about society and themselves, allowing new chances of integration by making possible interpersonal relationships without the strict necessity of standing in the same place.

By allowing remote communication in t-SL, PARLOMA makes peer-to-peer communication accessible. This is done by integrating haptic and robotic interfaces together with marker-less hand-tracking algorithms. PARLOMA allows local reception and reproduction of t-SL. The project poses the basis for the experience of a "telephone" for deaf-blind people.

In this paper, we discuss and evaluate a preliminary version of the system to prove that remote communication between deaf-blind signers is feasible. PARLOMA is designed to track a single signer's hand and to reproduce the performed hand-shape by using one robotic hand. It is worth noting here that such simplification does not affect the feasibility of the overall approach. In fact, we developed a general architecture that enables complete t-SL information transfer.

The full system is implemented on the Robot Operating System (ROS). ROS [11] is an open source, meta-operating system for robot software development, providing a collection of packages, software building tools and an

architecture for distributed inter-process and inter-machine communication. The building blocks of ROS-based applications are so-called "nodes". A node is a piece of code which implements a specific functionality. Nodes interact with each other by subscribing or publishing messages on specific ROS topics. Another interaction paradigm, which follows a request/reply model, is based on so-called "ROS services". The communication between nodes is based on the TCP network protocol.

We provide qualitative and quantitative analyses proving the effectiveness of the developed system. We also report preliminary tests in which selected hand-shapes from Italian SL (LIS) are recognized and sent remotely through ROS to a robotic hand that reproduces each of them to LIS signers. The results show that tactile communication is potentially highly effective, even when mediated by robotic interfaces. The main contributions of this paper are summarized below:

- By integrating computer vision techniques and robotic technologies, we conceive an assistive system that is able to transfer t-SL remotely. This is the first step to allow deaf-blind signers to communicate remotely.
- We design a robust, real-time, marker-less hand gesture recognition method that is able to track static hand-shapes robustly.
- We implement a preliminary working version of the system. Accordingly, we provide extensive qualitative and quantitative analyses to prove its feasibility, involving end-users in the experimental sessions.

This paper is organized as follows: Section 2 provides a detailed background on the targeted technologies; Section 3 discusses the theoretical approach and practical implementation of our solution; in Section 4 we present the results derived from our experiments and summarize the pipeline of the deaf-blind remote communication system we have developed; in Section 5, we discuss the results and propose future work to improve the system; finally, Section 6 concludes the paper.

2. Background

This paper is based on hand tracking, HGR and anthropomorphic robotic hands. In this section, we briefly summarize the state-of-the-art on these topics.

2.1 Hand Tracking

Object tracking techniques can be classified into two main classes: either *invasive* approaches, based on tools which are physically linked to the object (sensitized gloves [12] or markers [13]), or *non-invasive* approaches. The former are usually fast and computationally light, but often very expensive and cumbersome. The latter require more computational resources, but are generally based on low-cost technologies, and moreover do not require a physical link to the object to be tracked. As such, the object is free to move and not entangled.

Non-invasive approaches can be classified according to the kind of input data they need (2D, 3D) and the output that they provide [14]. Obviously, as real-world life is embedded in a 3D universe, the best performances are obtained when 3D feature characterization is performed [15], since 3D information is usually both more informative and more robust [16]. Moreover, low-cost acquisition systems, such as RGB-D cameras, represent a powerful tool for projects that aim at developing cheap and affordable solutions.

Non-invasive approaches are classified into *partial tracking* and *full tracking*. Tracking is defined as partial when it deals with only some insights of the kinematics of the hand (e.g., fingertip positions, discarding the rest of the hand), while in full approaches the whole hand is tracked. Of course, full tracking approaches are more useful for HRI applications, but they require much greater computational resources [17]. Full tracking approaches can be further divided into *model-based* and *appearance-based* approaches.

Model-based approaches [18] formulate the pose estimation as an optimization problem that minimizes the discrepancy between 3D hand hypotheses and the actual observations [19]. This problem is usually solved using optimization algorithms. In [8], the authors adopt a 3D hand model consisting of a set of assembled geometric primitives, animated by a vector of 27 parameters. A stochastic optimization method, called "Particle Swarm Optimization" (PSO) [20], is used to estimate the 27 model parameters' values that minimize the distance between the model itself and the input.

Appearance-based approaches employ discriminative techniques based on feature extraction from input data. At run-time algorithm extracts features from the input and tries to map them to a known hand pose. Appearance-based approaches are often implemented using database-retrieval [21] or machine learning [22] techniques. Learning how to map features and hand poses is the most computational intensive phase. Since this task is performed offline, while only fast feature extraction is required at run-time, appearance-based approaches easily achieve real-time performances. The accuracy of these algorithms is strongly related to the kinds of features and to the quality of the training set (or database), particularly to its variety and capacity to cover the whole set of poses.

2.2 Gesture Recognition

Currently, HGR is one of the main important research areas of computer science. Researchers are investigating different approaches by using Machine Learning, Neural Networks, Fuzzy Systems, Evolutionary Computation or Probabilistic Reasoning. Surveys on HGR algorithms are

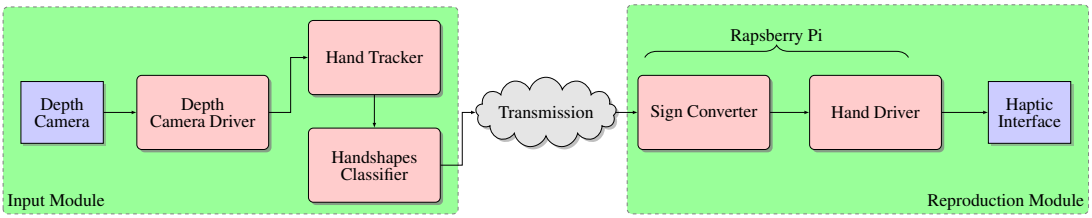

Figure 1. The Pipeline. The *Input Module* is in charge of performing hand-shape recognition using the depth image of the signer. It is composed of three ROS nodes and the depth camera itself. The recognized gesture is sent over the network to the *Reproduction Module*, which is composed of a robotic interface and two ROS nodes running on a Raspberry Pi.

given in [23, 2, 24]. The last two compare performances of various systems which have already been applied to human-machine interaction.

The accuracy of the system presented in [25] is about 90%, but it strongly depends upon the lighting conditions of the working environment. The average accuracy of the system presented in [26] is 80.77%, but this system only discriminates among simple gestures (e.g., showing a direction). [27] presents a system which is 90–95% accurate in recognizing open fingers while the success rate for closed fingers drops to 10–20% only. [28] claims around 90.45% accuracy, through hidden fingers could not be detected using this approach. The authors in [29] achieved 96% accuracy over 300 tests, but they developed an algorithm which does not meet the real-time constraint. The system presented in [30] achieves an accuracy rate of more than 95%, but it only works on six classes of gestures and it can only recognize bended fingers, no matters the degree of bending.

A remarkable work is presented in [31]. This paper focuses on building a robust part-based hand gesture recognition system using a Kinect sensor. The authors propose a method to handle the noisy hand shapes obtained from the depth camera. They propose a novel distance metrics, namely the Finger-Earth Mover's Distance, to measure the dissimilarity between hand shapes. Experimental results demonstrate that such a HGR system is quite accurate (a 93.2% mean accuracy is reported), but it works only on a 10-gesture dataset and it is quite slow (it achieves a 13 fps operating frequency). In [32], the authors only use RGB information, thus achieving different results in normal, brighter and darker conditions. The average accuracy of the system is in any case higher than 90% on discriminating hand poses for the 10 one-digit numbers (from 0 to 9). The operating frequency of the system is not reported.

2.3 Anthropomorphic Haptic Interfaces

Haptic devices elicit human perception through the sense of touch. Haptics therefore extends the communication channels for human-machine interaction in addition to the typical senses of vision and hearing.

Haptics includes wearable devices, such as gloves, and robotic devices, such as robotic arms and hands. With respect to robotic hands, despite the significant progress in recent decades in electronic integrated circuits and in applied computer science, challenges remain in increasing dexterity, robustness and efficiency, as well as in matching cost constraints [33]. Examples of dexterous robotic hands for humanoid robots are the *Awiwi* [34] and the *Shadow* hand [35].

Robotic hands designed to enable deaf-blind communication have already been proposed in the literature. The first attempt at creating a finger-spelling hand was patented in 1978 by the Southwest Research Institute (SWRI) [36]. Later, the Dexter hand was developed [37]. Dexter improved over the hand built by SWRI, but was extremely bulky and required compressed air to drive the pneumatic actuators. The whole hand had seven pneumatic actuators. Each finger was actuated by a single pneumatic actuator with a linear spring to provide some resistance and return. Both the thumb and the index finger had a second pneumatic actuator to perform complex letters.

The most successful design seems to be RALPH [38]. This hand was built in 1994 by the Rehabilitation Research and Development department of Veterans Affairs. RALPH fixed many of the problems of the Dexter hands, but it was still not robust or attractive. RALPH was only half a hand, as it only had fingers but no forearm and no wrist, which made it hard to read since it is in an unnatural position for the reader. In addition, it could perform only a limited subset of signs.

3. The Developed Solution

Our system is composed of a communication pipeline represented by three operations: (i) gesture acquisition and recognition *(front-end)*; (ii) gesture conversion and transmission; and (iii) gesture synthesis *(back-end)* (as depicted in Figure 1). The front-end and back-end are represented by two main sub-blocks, namely the *Input Module* and the *Reproduction Module*, respectively, while remote transmission is ensured by the ROS framework. The *Input Module* is connected to a depth camera. This module is able to identify gestures made by the human hand in front of the device. The *Reproduction Module* consists of a robotic hand and a controller which uses the information from the first module to control the robotic hand.

The ROS framework provides hardware abstraction; hence, the system is ready to control different robotic interfaces even in the case where multiple actuators are connected. In

addition, since ROS uses a distributed paradigm, remote communication is achieved in a simple manner. In the following, the entire pipeline will be explained.

3.1 Input Module

The *Input Module* is in charge of extrapolating 3D information from the depth map in order to understand the gesture performed by the user in real-time. It comprises three ROS nodes, namely the *Depth Camera Driver* node, the *Hand Tracker* node and the *Hand-shape Classifier* node.

The *Depth Camera Driver* node exposes the depth image stream from the depth camera as an ROS topic. In the proposed implementation, we use the *OpenNI* ROS driver for the *Asus Xtion* sensor [39].

3.1.1 Hand Tracker

The *Hand Tracker* node is a modified implementation of the algorithm proposed in [6], where the authors propose a full-DoF appearance-based hand tracking approach that uses a Random Forest (RF) classifier [40].

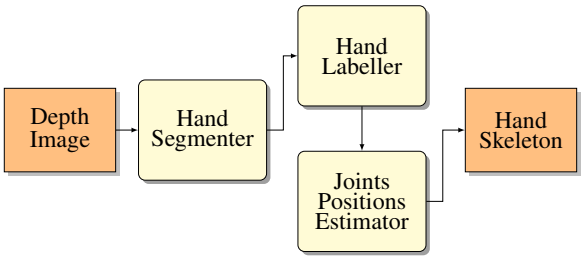

Figure 2. The *Hand Tracker* node. This node computes the hand joints' positions from depth images. It consists of three sequential tasks. The first one segments the hand (foreground) from the background. The second one classifies each pixel of the foreground to the hand region that it should belong to. The last one extrapolates the regions' centroids corresponding to the joints' positions.

This node extracts the hand skeleton and publishes the 3D position of each joint of the hand with respect to the camera reference frame. It accomplishes three main tasks performed sequentially on each incoming frame, as shown in Figure 2. In detail, the first task is accomplished by the *Hand Segmenter* block, which aims to distinguish the hand from the background.

Next, the *Hand Labeller* block executes an appearance-based approach to recognize different parts of the hand in order to isolate its joints. The *Joints Position Estimator* block approximates the joints' 3D positions, starting from the input depth measurements and the outcome of the first block. As in [40], in our approach an RF classifier [41] is used to label pixels of the depth image according to the region of the hand they belong to. Successively, each region is processed to find the position of its centre. At the end of the clustering process, the algorithm outputs the 3D position of each joint of the hand.

In our approach, we perform a per-pixel classification, where each pixel $\mathbf{x}$ of the hand is described using the following feature:

$$\mathcal{F}(\mathbf{x}) = \left\{ F_{\mathbf{u},\mathbf{v}}(I,\mathbf{x}), \|\mathbf{u}\| < R, \|\mathbf{v}\| < R \right\}, \quad (1)$$

where $I(\cdot)$ represents the depth value of the image at a given point, $\mathbf{u}$,$\mathbf{v}$ are two offsets limited to a finite R length, and the function $F_{u,v}(I,\mathbf{x})$ is defined as:

$$F_{\mathbf{u},\mathbf{v}}(I,\mathbf{x}) = I\left(x + \frac{\mathbf{u}}{I(\mathbf{x})}\right) - I\left(x + \frac{\mathbf{v}}{I(\mathbf{x})}\right).$$

This feature succeeds very quickly in discriminating hand parts [6]. Hand joints can be estimated by labelled segmented depth-maps using the Mean Shift (MS) algorithm [42]. In addition, the MS local mode-finding algorithm (as in [22]) reduces the risk of outliers, which might have a significant effect on the computation of the joints. By implementing MS, we obtain a more reliable and coherent estimation of the joints set $\mathcal{S}$.

Our labelling algorithm can recognize 22 different parts of the hand, namely the palm, the wrist and four joints for each of the fingers.

The joints' positions are approximated applying the MS clustering algorithm on the hand sub-parts. This approach shows promising results: experiments with real-world depth map images demonstrate that it can properly label most parts of the hand in real-time without requiring excessive computational resources.

Note that (1) is not invariant to rotations, while on the other hand it is invariant to distance and 3D translations (thanks to the normalization factor $I(\mathbf{x})$). As such, it is necessary to build a wide training set containing many instances of the same gesture captured from different points of view, according to [43]. For this reason, we have also investigated ways to effectively and automatically build comprehensive large training sets. Since manually building a dataset is a time-consuming and error-prone process, we also developed a tool that is able to create a synthetic training set. Such a system is based on the 3D model of a human hand shown in Figure 3. Essentially, the tool is able to generate intermediate hand-shapes from a small set of hand-shapes defined by the user. Next, all the hand-shapes are used to build the synthetic training using a model of the depth camera.

Table 1 summarizes the main learning parameters describing the RF used in our implementation. The proposed values have been experimentally evaluated. Each tree is trained with 2,000 pixels randomly sampled from each training image. Offset vectors $\mathbf{u}$ and $\mathbf{v}$ from (1) are sampled uniformly within a radius of 30 pixels.

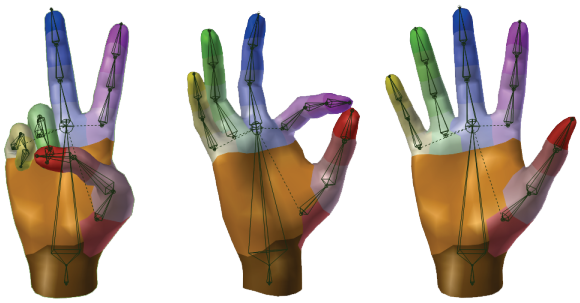

Figure 3. 3D model used to generate the synthetic training set. Our solution is able to generate intermediate hand-shapes from a small set of user-defined hand-shapes, in order to obtain a very large training set.

Parameter	Value
R	30 pixel
Threshold	10
Sample pixels per image	2000
Trees in the forest	3
Depth of each tree	18

Table 1. The optimal values that we propose to train the RF classifier

3.1.2 Hand-shape Classifier

The *Hand-shape Classifier* node is devoted to classify the current hand-shape according to the gesture that the user is making. This second classification is not performed directly on the array of joints' positions, but rather on a pattern containing the joint-to-joint Euclidean distances for all pairs of joints of the hand:

$$\mathcal{P}(\mathcal{S}) = \left\{ d_{k,l} = \left\| \mathbf{j}_k - \mathbf{j}_l \right\|, \forall \mathbf{j}_k, \mathbf{j}_l \in \mathcal{S}, k < l \right\}, \quad (2)$$

where $\mathbf{j}_{k/j}$ describes generic joints of the hand skeleton $\mathcal{S}$, while k and l are generic indexes within $\mathcal{S}$. The proposed pattern ensures invariance to rotation and translation in space.

Given $\mathcal{P}(\mathcal{S})$, another RF is used to evaluate the probability for a pose to actually reproduce one of the accepted gestures (i.e., hand-shapes shown in 3). The gesture is valid if the posterior probability associated with that hand-shape is above a p_{th} threshold and if it is recognized over N consecutive frames. In our experiments, p_{th} =0.3 and N =5.

Finally, the hand-shape is encapsulated in a network message that is sent to the reproduction device. Secure and lossless remote communication are guaranteed by the ROS (secure socket layer SSL).

3.2 Reproduction Module

The *Reproduction Module* is in charge of reproducing the recognized hand-shape by using the haptic interface, so that a t-SL-proficient receiver can understand the transmitted messages. This module is composed by two main ROS nodes, namely the *Sign Converter* node and the *Hand Driver* node. In the developed solution, this module runs on a low-cost computer (the Raspberry Pi). This architecture allows the system to have different output devices connected, enabling one-to-many communication among users.

3.2.1 Sign Converter

The *Sign Converter* node is in charge of converting recognized gestures in generic robotic hand poses (i.e., poses are not generated having in mind a specific robotic hand). This is achieved through an offline built dictionary that associates the gestures with a corresponding hand skeleton, called $\mathcal{S}^*$. The dictionary is populated using the same tool developed to populate the training set (see Section 3.1.1). After conversion, the skeleton $\mathcal{S}^*$ is finally sent to the *Hand Driver*.

Note that $\mathcal{S}^*$ is a static representation of the hand skeleton $\mathcal{S}$ computed by the *Hand Tracker* node. With this architecture, we control the haptic interface with sharp commands, avoiding the effects of the noise affecting $\mathcal{S}$.

3.2.2 Hand Driver

Finally, the *Hand Driver* node is devoted to control the robotic hand with specific commands depending upon the robotic hand employed. It receives as input the skeleton $\mathcal{S}^*$ and performs specific algorithms of inverse kinematics and collision avoidance. Collision avoidance is of extreme importance when it is necessary to switch between poses that are really different, requiring a majority of fingers (or all of them) to move.

The *Hand Driver* is also in charge of performing the hardware abstraction of the robotic interface, and is the only node in the developed architecture that has to be changed when different haptic interfaces are used. The ROS architecture also enables one-to-many communication. In this case, multiple *Hand Driver* nodes need to be instantiated according to the different simultaneously-operated actuators.

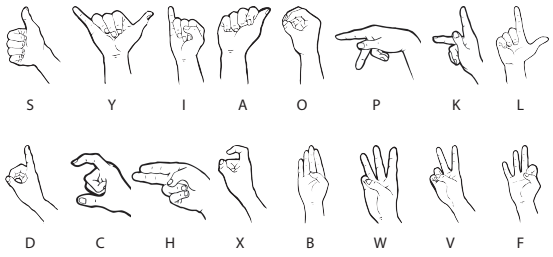

Figure 4. Hand-shapes from the LIS alphabet used in the experiments

3.3 Haptic Interface

The *Haptic Interface* used is a right-handed version of the CyberHand [44]. It consists of four under-actuated anthropomorphic fingers and a thumb, and it is actuated by six DC motors. Five of them, located remotely, control finger flexion/extension. One motor, housed inside the palm, drives the thumb abduction/adduction. The hand is comparable in size to the adult human hand, and the remote actuators are assembled in an experimental platform which mimics the shape of the human forearm. The remote actuators act on their respective fingers using tendons and a Bowden cable transmission. Active flexion is achieved as follows: when a tendon is pulled, the phalanxes flex synchronously, replicating the idle motion (i.e., free space motion) of a human finger. When a tendon is released, torsion springs located within the joints extend the fingers. The hand includes encoders on each motor and an electronic controller that implements position control by receiving commands sent over a serial bus.

4. Experimental Results

Experimental setup. On the left, a signer performing hand-shapes in front of the RGB-D Camera. Information is elaborated and sent over the net to a robot hand. On the right, the receiver is able to understand the sign performed by the robot hand using tactile sensing.

This section presents the results of the various tests that we performed. Firstly, we test the recognition module (Section 4.1), i.e., the ability of the system to recognize hand-shapes; secondly, we test the transmission efficiency of the whole system (Section 4.2) by measuring the loss of information during the recognition and reproduction phases in a test-case scenario; finally, an experimental session performed with a deaf-blind subject is presented (Section 4.3) to assess the overall usability of the system.

These experiments focus on hand-shapes, i.e., static configurations of the hand. We chose finger-spelling as the hand-shapes source. Finger-spelling consists of spelling complex language words using the manual alphabet. While the usage of finger-spelling in the SL lexicon may be limited, it is definitely larger in t-SL. The decision for using finger-spelling in this preliminary phase is due to: i) the fact that the robotic hand was originally conceived for prosthetic applications [45], with a constrained number of actuators following an under-actuation scheme, and therefore cannot perform the full set of gestures; ii) our interest in getting feedback on this specific component of the sign, namely the hand-shape. We use only 16 hand-shapes (Figure 4), which are the hand-shapes of the LIS manual alphabet that the robot hand can correctly reproduce. Of course, alternative (even meaningless) hand-shapes could have been used.

4.1 Input Module Evaluation

The tests evaluate and quantify the system performance in recognizing hand-shapes (the first step of the pipeline). Sixteen subjects were recruited (10 men, six women, mean age 25 years, range ± four years). None of the subjects had any expertise with LIS. Indeed, the subjects were chosen in order to check the accuracy of the system even in the case of potential beginner users of the final product. The experiment consists in the production of isolated hand-shapes corresponding to the manual alphabet letters. Subjects were sat at a table in front of a laptop (a Macbook PRO, late 2011, mounting an Intel Core i7@2.7GHz CPU, 4GB of RAM, an Intel HD Graphics 3000 512MB GPU) and a depth camera (Asus Xtion PRO), and had to repeat each of the hand-shapes appearing on the laptop's monitor (as shown in Figure 5). Each subject was 50–55 cm away from the camera lens. For each pose, the system recorded 100 depth maps (at a rate of 30 fps). This part of the experiment lasted about three minutes per person (six or seven seconds per hand-shape per person). In total, 40 thousand depth maps were collected. Each pose was then linked with the corresponding LIS alphabet letter.

With respect to the 40 thousand depth maps acquired, we performed 10 leave-one-out cross-validations to investigate the accuracy of the *Input Module*. For each

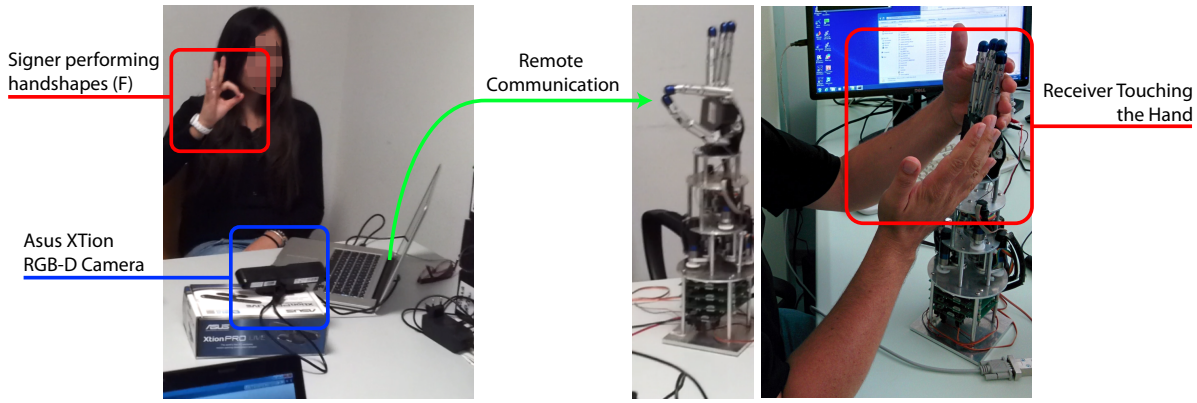

Figure 5. Experimental setup. On the left, a signer performing hand-shapes in front of the RGB-D Camera. Information is elaborated and sent over the net to a robot hand. On the right, the receiver is able to understand the sign performed by the robot hand using tactile sensing.

	S	Y	I	A	O	P	K	L	D	C	H	X	B	W	V	F
S	90	4	0	6	0	0	0	0	0	0	0	0	0	0	0	0
Y	21	44	0	32	0	0	0	0	0	0	0	0	0	3	0	0
I	12	3	66	16	0	0	0	0	3	0	0	0	0	0	0	0
A	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0	0
O	0	0	0	0	100	0	0	0	0	0	0	0	0	0	0	0
P	0	0	0	75	8	9	0	0	0	0	0	8	0	0	0	0
K	0	0	0	0	0	39	36	17	4	0	0	4	0	0	0	0
L	7	0	0	0	0	0	0	93	0	0	0	0	0	0	0	0
D	0	0	0	3	0	0	3	10	84	0	0	0	0	0	0	0
C	0	0	0	0	0	0	0	3	0	88	0	0	3	0	0	6
H	0	0	0	0	0	0	0	6	6	0	76	6	0	0	6	0
X	5	0	0	0	0	0	0	0	15	0	0	75	0	5	0	0
B	0	0	0	0	0	0	0	0	0	0	0	0	87	13	0	0
W	0	0	0	0	0	4	0	0	0	0	0	0	8	79	9	0
V	0	0	0	0	0	0	3	0	0	0	0	0	0	9	88	0
F	0	0	0	0	0	0	0	0	0	0	0	0	0	2	1	97

Figure 6. Confusion matrix of the *Input Module*. On the Y axis, the ground truth, while on the X axis are the gestures recognized by the module. Similar hand-shapes are presented in adjacent cells.

validation, depth maps from 15 subjects were used for training the RF classifier from the *Hand-shape Classifier* node (the training procedure was similar to that in [43]), while data from one random subject are used for testing it. The results are summarized in Figure 6, which shows that the average accuracy of the *Input Module* system with respect to the ground truth set is 74%, with a confidence interval of ±2%.

It should be noticed that noisy data from the sensor are the main source of errors in the case of the gesture "P", which on 75% of occasions was recognized as "A". "P" hand-shapes differ from the latter only because the index is pointing at the camera (instead of being closed near the palm, see Figure 4); however, very often noise totally covers the index finger in such a configuration, such that is impossible for our algorithms to discriminate between these two gestures. Different is the case of the gesture "K", which on 39% of occasions was recognized as "P" and 17% of the times as "L". These errors are due to the intrinsic properties of (2), which cannot discriminate enough different poses in which the relative joint distances do not vary. In fact, note that "K" seems like a "P" that is rotated clockwise by 90 degrees, and an "L" in which the thumb and middle-finger bending are exchanged.

The average accuracy obtained by our system is comparable with [46], especially considering that we track the human hand frame by frame. Our algorithms do not require extra computation to estimate the arm or even the full body pose. However, other state-of-the-art approaches perform slightly better (e.g., [47, 48]). This is typically due to our per-pixel classification (within the *Hand Tracker* node), which is not robust enough yet, and to the depth camera that we use, which produces very noisy data. In any case, our experiments confirm that the average accuracy achieved by our approach is sufficient to effectively track the hand, even with respect to challenging backgrounds, and to discriminate among valid hand-shapes for SL-based interactions (even if they are similar).

4.2 Transmission Evaluation

Another test section involved the entire pipeline. Seven subjects with some expertise in LIS but no previous knowledge about the project participated in the test (four men, three women, mean age 30 years, range ± four years) together with an LIS signer (female, 24 years old). Since the status of the project is still in its preliminary stages, the test was not assessed by deaf-blind signers.

Each experiment was performed as hereby described and depicted in Figure 5. The results along the whole pipeline are illustrated in Figure 7. A list of 125 hand-shapes (LoS) to be reproduced (up to 10 repetitions for each hand-shape) was randomly generated at run-time. The LoS appeared on a monitor close to the LIS expert signer. She reproduced with the right hand each hand-shape in front of the depth camera. Also in this case, signs were made from a distance of 50–55 cm from the camera lens. To simulate a real-use case scenario, the LIS expert was placed in unfavourable lighting conditions, and people were allowed to pass behind her. Recognized hand-shapes were saved for successive processing in a dedicated list (SRI).

In a different room in the same building (so that remote communication via the Internet could be simulated), a Raspberry Pi received the ROS messages, decoding them and generating the commands needed to control the hand. The task for the subjects (who were in this other room), was to visually recognize the hand-shape actually performed by the robotic hand. Answers were recorded and saved in a separate list (SRS). The subjects assessed the task individually. The hand reproduced each hand-shape for five seconds. After this period of time, it returned to the rest

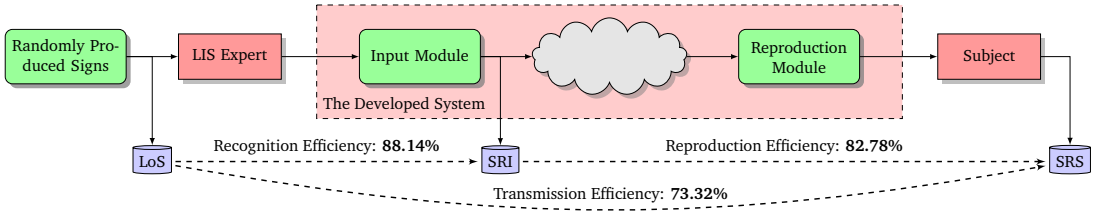

Figure 7. Sign recognition and reproduction efficiency along the pipeline of the experimental apparatus

position (with all the fingers opened) before the next sign was reproduced.

During the experiments, all possible feedback was collected and stored: the list of hand-shapes to be reproduced and the list of signs recognized by the *Input Module* and by the volunteer, depth maps from the camera, data from the network, and feedback about the joint positions of the robotic hand. Network latency during all the experiments was, on average, below half a second, and was never higher than one second.

The experiment lasted for approximately 12 minutes per person. At the end, the experimenters asked the volunteers for their comments, especially regarding any major difficulties that they experienced and their opinions on the usability of the hand for their purposes.

The collected results (summarized in Figure 7) demonstrate the feasibility of our system. In total, 875 hand-shapes were performed by the hand (125 hand-shapes per seven volunteers). The average accuracy of the acquisition module was 88.14%; however, few errors reported for the *Input Module* were actually caused by the expert in LIS performing the wrong hand-shape (e.g., performing gesture "A" while the experiment was asking for "I"). No errors where registered for the transmission and conversion systems. On 82.78% of occasions, the signs were correctly recognized by the volunteers and the total accuracy of the system in transmission was 73.32%.

In Figure 7, *Recognition Efficiency* refers to the percentage of hand-shapes correctly recognized by the *Input Module*. This comparison was to evaluate the effectiveness of the recognition module. Here, errors were due to classification errors, noisy data from the depth camera and finger occlusions occasionally deceiving the recognition algorithm.

Reproduction Efficiency refers to the percentage of hand-shapes correctly recognized by the subjects. Not surprisingly, we found that most of the time the volunteers simply confused a hand-shape for a similar one. As a consequence of the constrained number of degrees of freedom of the robotic hand in comparison to the human hand, many subjects reported minor difficulties in discriminating similar signs when different degrees bending would have been essential for the signs' discrimination. The most frequent recognition errors were the following:

- 17 times, the letter "P" was confused with "X". In particular, the hand-shapes for "P" and "X" are very similar; in both cases, all the fingers except for the index finger were closed. Performing "P" requires bending just the metacarpophalangeal (MCP) joint of the index (the others are opened), while when performing "X" all the index finger's joints are bent: however, as a consequence of the under-actuation scheme, the robot hand cannot move different joints of the same finger independently.
- 12 times, the letter "O" was confused with "C". The hand-shapes for the "O" and "C" signs are very similar: both include all five fingers becoming close to one another; the only difference between them is that the fingers are closer to each other while performing "O". The robotic hand that we used for the experiments could not cover all of the intermediate positions from "finger open" to "finger closed", as it has one degree of freedom per underactuated finger, and so it cannot properly emphasize the difference between "C" and "O".

Finally, *Transmission Efficiency* measures the efficiency of the whole experimental apparatus. The accuracy of the entire pipeline from the hand-shape produced by the signer to the visual recognition by the subjects is significantly high ($\chi^2 = 215.764$, $df = 1$, $p < 0.001$).

4.3 Reproduction Module Evaluation

A deaf-blind signer member of the Lega del Filo d'Oro, Francesco Ardizzino (male, 64 years old) volunteered to test the robot hands during the design process. His experience was then reported to the association panel and the PARLOMA project received official approval from the association.

In order to test the *Reproduction Module*, we checked whether a LIS-t deaf-blind signer was able to: i) recognize isolated hand-shapes produced by an anthropomorphic robotic hand via tactile exploration (i.e., the natural setting for t-SL communication); ii) to read-off a finger-spelled word from a sequence of hand-shapes.

An Italian LIS-t interpreter was present during the entire phase of testing (two hours and 30 minutes, including some breaks between stages and major tasks). The interpreter explained each task to the subject, who gave consent to participating in the experiment and authorized us to record the whole section.

Before assessing the accuracy in recognizing hand-shapes from the robotic hand, we made sure that the signer was able to assess the same tasks when performed by a human as a baseline. As such, we asked the interpreter to produce a few hand-shapes in isolation and then a few finger-spelled words. All of them were recognized by our subject.

The whole section is structured as a sequence of stages. The first is concerned with isolated digit and hand-shape recognition. The hand performs a set of poses (hand-shapes) commanded by an experimenter using an Acer Aspire 5810TZ laptop running a custom graphics interface program. Since, in this case, we are not interested in testing the entire pipeline, we ensured that the robotic hand correctly performed the hand-shape. The subject was asked to explore the robotic hand and then report which sign he comprehended. Tactile exploration is very quick and, especially in the first stages, it affected all parts of the hand. This is somehow different from what normally happens during human t-SL communication, where the exploration

is more holistic. We attribute this to the fact that the subject was experiencing a robotic hand for the first time.

The first task was to recognize digits from one to five (including some variants, such as "1" made by extending the index finger or the thumb). The task was easily assessed and the digits were recognized appropriately. In this phase, we noticed that the subject was not actually counting the finger but only assessing the global hand shape of the hand, confirming that part-by-part exploration is related to the novelty of the interaction. The only problematic digit to recognize was "5". In total, the volunteer was asked to recognize eight digits, and he failed only once (success rate 88%).

The second task involved the recognition of hand-shapes corresponding to the letters of the manual alphabet. All the letters were presented randomly. Only the letters "G", "J", "Z" and "R" were excluded, the first three because they require dynamic gestures, and the last one because it involves the crossing of the index and middle fingers (an option not available to the hand). When needed, the hand was manually oriented, such as in the case of the letters "M", "N" or "Q". The subject recognized most of the hand-shapes very quickly. The most problematic hand-shapes were "Q", which was often mistaken for "P"; "E" and "O" were often mistaken for "A". Following each letter, the LIS interpreter would tell the user whether his answer corresponded to the real hand-shape. In total, the subject was asked to recognize 19 digits, and he failed twice (success rate 89%).

The second stage concerned finger-spelled words. For this task, we asked the subject to recognize sequences of hand-shapes corresponding to the letters of the manual alphabet, and retained them until he could read-off an Italian word. We were interested not only in "simple" words, made of letters that are easily recognized (e.g., C-A-N-I = "dogs", S-L-I-T-T-A = "sledge"), but especially in more complex words containing problematic letters not easily recognized in the previous step (e.g., A-C-Q-U-A = "water", A-I-U-O-L-E = "flowerbeds"). As soon as the subject was able to read-off the Italian word, he was asked to produce the corresponding LIS sign. No errors were registered in this final experiment, showing that the integration with context improves a correct hand-shape recognition.

5. Discussion

This section is devoted to a brief discussion of the experimental results reported in the previous section and presents planned future work devoted to improving the entire system.

We implemented a *Hand-shape Classifier* node (see Section 3.1.2), which is needed because the first RF returns a very noisy estimation of the hand joints' positions. Clearly, this prevents the input module from directly controlling the robotic hand by using the estimated joints positions. One source of noise is the input camera itself. Asus Xtion is a low-cost device, developed for tracking the whole human body from afar and not specifically for tracking small objects like human hands at short-range. PARLOMA shows that it is also possible to achieve satisfactory performances from such a low-cost device.

Particular effort will be devoted to improving the performances of the tracking algorithm (Section 3.1.1). In addition, as Section 4.1 points out, improvements are still required in the *Hand-shape Classifier* node to extend our classification pattern so as to also consider the absolute orientation of the hand-shapes for the classification task (Equation 2). This is necessary for discriminating between poses that have the same joints' configuration but different orientations in space.

Moreover, our project requires the development of a haptic interface specifically targeted for t-SL communication. The hand used in the proposed implementation is an anthropomorphic hand specifically designed for prosthetics (especially for grasping). We plan to develop a low-cost robotic hand specifically targeted to mimic the human hand's high level of dexterity. This hand will come with more degrees of freedom and faster motors, and will be 3D-printed for faster prototyping purposes. A higher hand dexterity will improve PARLOMA performances, as it will help receivers to better recognize letters. Further developments will take feedback collected during the experiments proposed in this paper strongly into account (see Section 4.2 and Section 4.3). This new, low-cost robot hand will then be used as a starting point for developing a complete anthropomorphic robot able to completely mimic the dexterity of the human body, which is needed to emulate t-SL communication in its entire complexity. A complete robot will be able to fully satisfy the requirements needed by our main target, i.e., a complete and full transmission of messages coded in t-SL. We aim to develop a genuinely low-cost haptic interface to make our system affordable for subsidized healthcare programmes.

Finally, we are reconsidering the global architecture of the proposed system. To date, there is a strong asymmetry in PARLOMA, as a low-cost and simple credit card-sized computer is used to control the output device, while a costly one with a fast and powerful GPU is required to run the complex algorithms behind the input module. For this reason, we are interested in developing a novel architecture based on the *Cloud Robotics* paradigm [49], where the most complex algorithms run on a remote server and the input/output intelligence is devoted solely to acquiring images, controlling the haptic interface and guaranteeing remote communication.

6. Conclusion

In this paper we present PARLOMA, a system that allows non-invasive remote control of a robotic hand by means of gesture recognition techniques, using a single low-cost depth camera as only input. This system is able to recognize

human hand poses in real time and to send them over the Internet through the ROS framework. Poses are received by a remote computer that can control a robotic hand, that mimics human hand poses. The system is at its early stage of development. Yet, it already enables Deaf-blind people to remotely communicate resorting on the LIS manual alphabet. To the best of our knowledge, this is the first system offering this capability. Moreover, PARLOMA is developed in strong collaboration with Deaf-blind associations and is thought to be intuitive to use and effective. A focus group composed of non-disabled persons proficient in t-LS and a Deaf-blind signer have been actively involved during all the design tasks. Through all the paper we demonstrate the feasibility of PARLOMA, together with open issues that will be addressed by our future work.

Although PARLOMA currently works with static handshapes only, the system is expected to send real t-SL messages (including the dynamic component). Moreover, it intrinsically support one-to-many communication among Deaf-blind people.

Our experimental results show that the system works correctly. Section 4.1 discusses the performances of the input system (accuracy 88.14%) while Section 4.2 evaluates the whole pipeline (accuracy 73.32%).

PARLOMA has the potential to completely capture and transmit any t-SL based message. t-SL messages are coded through handshapes and their evolution in time and space. The developed acquisition system is already able to capture entirely the first information and is fast enough to guarantee an appropriate sampling time to let the robotic hand mimicking time evolution. In this way, we are able to capture both verbal information (signs) but also non verbal aspects, such as inflections and tones of communication. In addition, the very low quantity of information extrapolated from the input device is enough compact to be sent over the net using the standard technology without introducing delays in communication.

We are investigating alternative applications for the technology developed and presented in this paper. For instance, the proposed architecture can be used to control different robotic devices, such as an hand exoskeleton, that can guide users in post-stroke rehabilitation to replicate correctly movements that a physiotherapist is performing in front of a camera. The effectiveness of tele-rehabilitation is well demonstrated in literature.

In conclusion, PARLOMA is the first step through a complete remote communication system for Deaf-blind people, and this paper shows that the present architecture has the potential to evolve quickly in a fully working system able to work as a telephone for t-SL communication. In fact, as the telephone is able to capture, remotely send and reproduce (without interoperation) the mean of the vocal languages, that are sounds, our system has the potential to do the same with the mean of the t-SL information, i.e., hands and arms movements in space and time.

7. Acknowledgements

The authors would like to thank Professor Paolo Prinetto and Professor Basilio Bona from Politecnico di Torino for their valuable help and support.

We also thank Lega del Filo d'Oro, Francesco Ardizzino (our Deafblind consultant) and Alessandra Checchetto (the LIS-t interpreter) for the help and the support they gave us.

This research was partially supported by the "Smart Cities and Social Innovation Under 30" programme of the Italian Ministry of Research and Universities through the PARLOMA Project (SIN_00132) and by the Italian Ministry of Health, Ricerca Finalizzata 2009 - VRehab Project (Grant RF-2009-1472190).

Part of the research was supported by Telecom Italia S.p.A.

Part of the research leading to these results received funding from the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013) / ERC Grant Agreement Nº 324115–FRONTSEM (PI: Schlenker). Part of this research was conducted at Institut d'Etudes Cognitives (ENS), which is supported by grants ANR-10-IDEX-0001-02 PSL* and ANR-10-LABX-0087 IEC.

Part of the research leading to these results has received funding from the Agence Nationale de la Recherche (grants ANR-10-LABX-0087 IEC and ANR-10-IDEX-0001-02 PSL).

8. References

[1] Flávio Garcia Pereira, Raquel Frizera Vassallo, and Evandro Ottoni Teatini Salles. Human-robot interaction and cooperation through people detection and gesture recognition. *Journal of Control, Automation and Electrical Systems*, 24(3):187–198, 2013.

[2] Ankit Chaudhary, Jagdish Lal Raheja, Karen Das, and Sonia Raheja. Intelligent approaches to interact with machines using hand gesture recognition in natural way: A survey. *CoRR*, abs/1303.2292, 2013.

[3] Matthieu Bray, Esther Koller-Meier, and Luc Van Gool. Smart particle filtering for 3d hand tracking. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 675–680. IEEE, 2004.

[4] Srinath Sridhar, Antti Oulasvirta, and Christian Theobalt. Interactive markerless articulated hand motion tracking using rgb and depth data. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 2456–2463. IEEE, 2013.

[5] Chen Qian, Xiao Sun, Yichen Wei, Xiaoou Tang, and Jian Sun. Realtime and robust hand tracking from depth. In *Computer Vision and Pattern Recognition*

(CVPR), 2014 IEEE Conference on, pages 1106–1113, June 2014.

[6] Cem Keskin, Furkan Kraç, Yunus Emre Kara, and Lale Akarun. Real time hand pose estimation using depth sensors. In *Consumer Depth Cameras for Computer Vision*, pages 119–137. Springer, 2013.

[7] Robert Wang, Sylvain Paris, and Jovan Popovic. 6d hands: markerless hand-tracking for computer aided design. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 549–558. ACM, 2011.

[8] Iason Oikonomidis, Nikolaos Kyriazis, and Antonis A Argyros. Efficient model-based 3d tracking of hand articulations using kinect. In *BMVC*, pages 1–11, 2011.

[9] Wendy Sandler and Diane Lillo-Martin. *Sign language and linguistic universals*. Cambridge University Press, 2006.

[10] Alessandra Checchetto, Carlo Cecchetto, Carlo Geraci, Maria Teresa Guasti, and Alessandro Zucchi. Una varietà molto speciale: La list (lingua dei segni italiana tattile). *Grammatica, lessico e dimensioni di variazione nella LIS*, page 207, 2012.

[11] Morgan Quigley, Ken Conley, Brian Gerkey, Josh Faust, Tully Foote, Jeremy Leibs, Rob Wheeler, and Andrew Y Ng. Ros: an open-source robot operating system. In *ICRA workshop on open source software*, volume 3.2, page 5, 2009.

[12] Federico Lorussi, Enzo Pasquale Scilingo, Mario Tesconi, Alessandro Tognetti, and Danilo De Rossi. Strain sensing fabric for hand posture and gesture monitoring. *Information Technology in Biomedicine, IEEE Transactions on*, 9(3):372–381, 2005.

[13] Robert Y Wang and Jovan Popovic. Real-time hand-tracking with a color glove. In *ACM Transactions on Graphics (TOG)*, volume 28, page 63. ACM, 2009.

[14] James M Rehg and Takeo Kanade. Digiteyes: Vision-based hand tracking for human-computer interaction. In *Motion of Non-Rigid and Articulated Objects, 1994., Proceedings of the 1994 IEEE Workshop on*, pages 16–22. IEEE, 1994.

[15] Kevin Bowyer, Kyong Chang, and Patrick Flynn. A survey of approaches and challenges in 3d and multi-modal 3d+2d face recognition. *Computer Vision and Image Understanding*, 101(1):1 – 15, 2006.

[16] Luis Goncalves, Enrico Di Bernardo, Enrico Ursella, and Pietro Perona. Monocular tracking of the human arm in 3d. In *Computer Vision, 1995. Proceedings., Fifth International Conference on*, pages 764–770. IEEE, 1995.

[17] Ali Erol, George Bebis, Mircea Nicolescu, Richard D Boyle, and Xander Twombly. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*, 108(1):52–73, 2007.

[18] Dariu M Gavrila and Larry S Davis. 3-d model-based tracking of humans in action: A multi-view approach. In *Computer Vision and Pattern Recognition, 1996. Proceedings CVPR'96, 1996 IEEE Computer Society Conference on*, pages 73–80. IEEE, 1996.

[19] Jungong Han, Ling Shao, Dong Xu, and Jamie Shotton. Enhanced computer vision with microsoft kinect sensor: A review. *IEEE Trans. Cybernetics*, 43(5), October 2013.

[20] James Kennedy and Russell Eberhart. Particle swarm optimization. In *Proceedings of 1995 IEEE International Conference on Neural Networks*, pages 1942–1948, 1995.

[21] Vassilis Athitsos and Stan Sclaroff. Estimating 3d hand pose from a cluttered image. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume 2, pages II–432. IEEE, 2003.

[22] Jamie Shotton, Toby Sharp, Alex Kipman, Andrew Fitzgibbon, Mark Finocchio, Andrew Blake, Mat Cook, and Richard Moore. Real-time human pose recognition in parts from single depth images. *Communications of the ACM*, 56(1):116–124, 2013.

[23] Sushmita Mitra and Tinku Acharya. Gesture recognition: A survey. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, 37(3):311–324, 2007.

[24] Vladimir I Pavlovic, Rajeev Sharma, and Thomas S. Huang. Visual interpretation of hand gestures for human-computer interaction: A review. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 19(7):677–695, 1997.

[25] Jagdish Lal Raheja, Radhey Shyam, Umesh Kumar, and P Bhanu Prasad. Real-time robotic hand control using hand gestures. In *Machine Learning and Computing (ICMLC), 2010 Second International Conference on*, pages 12–16. IEEE, 2010.

[26] Heung-Il Suk, Bong-Kee Sin, and Seong-Whan Lee. Hand gesture recognition based on dynamic bayesian network framework. *Pattern Recognition*, 43(9):3059–3072, 2010.

[27] Dung Duc Nguyen, Thien Cong Pham, and Jae Wook Jeon. Fingertip detection with morphology and geometric calculation. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 1460–1465. IEEE, 2009.

[28] Ekaterini Stergiopoulou and Nikos Papamarkos. Hand gesture recognition using a neural network

shape fitting technique. *Engineering Applications of Artificial Intelligence*, 22(8):1141–1158, 2009.

[29] Alexandra Stefan, Vassilis Athitsos, Jonathan Alon, and Stan Sclaroff. Translation and scale-invariant gesture recognition in complex scenes. In *Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments*, page 7. ACM, 2008.

[30] Jong Shill Lee, Young Joo Lee, Eung Hyuk Lee, and Seung Hong Hong. Hand region extraction and gesture recognition from video stream with complex background through entropy analysis. In *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, volume 1, pages 1513–1516. IEEE, 2004.

[31] Ren Zhou, Yuan Junsong, Meng Jingjing, and Zhang Zhengyou. Robust part-based hand gesture recognition using kinect sensor. *Multimedia, IEEE Transactions on*, 15(5):1110–1120, Aug 2013.

[32] Feng Kai-ping and Yuan Fang. Static hand gesture recognition based on hog characters and support vector machines. In *Instrumentation and Measurement, Sensor Network and Automation (IMSNA), 2013 2nd International Symposium on*, pages 936–938, Dec 2013.

[33] Marco Controzzi, Christian Cipriani, and Maria Chiara Carrozza. Design of artificial hands: A review. In *The Human Hand as an Inspiration for Robot Hand Development*, pages 219–246. Springer, 2014.

[34] Markus Grebenstein. The awiwi hand: An artificial hand for the dlr hand arm system. In *Approaching Human Performance*, pages 65–130. Springer, 2014.

[35] R Walkler. Developments in dextrous hands for advanced robotic applications. In *Proc. the Sixth Biannual World Automation Congress, Seville, Spain*, pages 123–128, 2004.

[36] Brian Fang, Colby Dixon, and Trevor Wong. Robotic fingerspelling hand for deaf-blind communication. 2012.

[37] A. Meade. Dexter–a finger-spelling hand for the deaf-blind. In *Robotics and Automation. Proceedings. 1987 IEEE International Conference on*, volume 4, pages 1192–1195, Mar 1987.

[38] David L Jaffe. Ralph, a fourth generation fingerspelling hand. *Rehabilitation Research and Development Center*, 1994.

[39] Krystof Litomisky. Consumer rgb-d cameras and their applications. Technical report, Tech. rep. University of California, 2012.

[40] Victor Francisco Rodriguez-Galiano, Bardan Ghimire, John Rogan, Mario Chica-Olmo, and Juan Pedro Rigol-Sanchez. An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67:93–104, 2012.

[41] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[42] Dorin Comaniciu and Peter Meer. Mean shift: A robust approach toward feature space analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(5):603–619, 2002.

[43] Cem Keskin, Furkan Kraç, Yunus Emre Kara, and Lale Akarun. Hand pose estimation and hand shape classification using multi-layered randomized decision forests. In *Computer Vision–ECCV 2012*, pages 852–863. Springer, 2012.

[44] Maria Chiara Carrozza, Giovanni Cappiello, Silvestro Micera, Benoni B Edin, Lucia Beccai, and Christian Cipriani. Design of a cybernetic hand for perception and action. *Biological cybernetics*, 95(6): 629–644, 2006.

[45] Stanisa Raspopovic, Marco Capogrosso, Francesco Maria Petrini, Marco Bonizzato, Jacopo Rigosa, Giovanni Di Pino, Jacopo Carpaneto, Marco Controzzi, Tim Boretius, Eduardo Fernandez, et al. Restoring natural sensory feedback in real-time bidirectional hand prostheses. *Science translational medicine*, 6(222):222ra19–222ra19, 2014.

[46] Zahoor Zafrulla, Helene Brashear, Thad Starner, Harley Hamilton, and Peter Presti. American sign language recognition with the kinect. In *Proceedings of the 13th international conference on multimodal interfaces*, pages 279–286. ACM, 2011.

[47] Danhang Tang, Tsz-Ho Yu, and Tae-Kyun Kim. Real-time articulated hand pose estimation using semi-supervised transductive regression forests. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3224–3231. IEEE, 2013.

[48] Chi Xu and Li Cheng. Efficient hand pose estimation from a single depth image. In *Computer Vision (ICCV), 2013 IEEE International Conference on*, pages 3456–3462. IEEE, 2013.

[49] Dominique Hunziker, Mohanarajah Gajamohan, Markus Waibel, and Raffaello D'Andrea. Rapyuta: The roboearth cloud engine. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 438–444. IEEE, 2013.